

## NEGATIVE ASSOCIATION, ORDERING AND CONVERGENCE OF RESAMPLING METHODS

BY MATHIEU GERBER\*, NICOLAS CHOPIN<sup>†,1</sup> AND NICK WHITELEY\*

*University of Bristol\* and CREST-ENSAE<sup>†</sup>*

We study convergence and convergence rates for resampling schemes. Our first main result is a general consistency theorem based on the notion of negative association, which is applied to establish the almost sure weak convergence of measures output from Kitagawa's [*J. Comput. Graph. Statist.* **5** (1996) 1–25] stratified resampling method. Carpenter, Ckiffird and Fearnhead's [*IEE Proc. Radar Sonar Navig.* **146** (1999) 2–7] systematic resampling method is similar in structure but can fail to converge depending on the order of the input samples. We introduce a new resampling algorithm based on a stochastic rounding technique of [In *42nd IEEE Symposium on Foundations of Computer Science (Las Vegas, NV, 2001)* (2001) 588–597 IEEE Computer Soc.], which shares some attractive properties of systematic resampling, but which exhibits negative association and, therefore, converges irrespective of the order of the input samples. We confirm a conjecture made by [*J. Comput. Graph. Statist.* **5** (1996) 1–25] that ordering input samples by their states in  $\mathbb{R}$  yields a faster rate of convergence; we establish that when particles are ordered using the Hilbert curve in  $\mathbb{R}^d$ , the variance of the resampling error is  $\mathcal{O}(N^{-(1+1/d)})$  under mild conditions, where  $N$  is the number of particles. We use these results to establish asymptotic properties of particle algorithms based on resampling schemes that differ from multinomial resampling.

**1. Introduction.** A resampling scheme is a randomized procedure that takes as input random samples  $X^n$  with nonnegative weights  $W^n \geq 0$ ,  $n = 1, \dots, N$ , such that  $\sum_{n=1}^N W^n = 1$ , and returns as an output resampled variables  $X^{A^n}$ , where  $A^n$  is a random index in  $\{1, \dots, N\}$ , such that, in some sense,

$$(1) \quad \frac{1}{N} \sum_{n=1}^N \delta(X^{A^n}) \approx \sum_{n=1}^N W^n \delta(X^n).$$

Here,  $\delta(x)$  denotes the Dirac measure at point  $x$  (this slightly unconventional notation will make our equations more readable).

Resampling appears in various statistical procedures. The present work is primarily motivated by resampling within Sequential Monte Carlo methods, also known as particle filters [Doucet, de Freitas and Gordon (2001)]. Particle filters

---

Received August 2017; revised June 2018.

<sup>1</sup>Supported in part by Labex Ecodec (anr-11-labx-0047).

*MSC2010 subject classifications.* Primary 62G09, 62M20; secondary 60G35.

*Key words and phrases.* Negative association, resampling methods, particle filtering.

approximate recursively a sequence of probability distributions by propagating  $N$  “particles” through weighting, resampling and mutation steps. The resampling steps play a crucial role in stabilizing the Monte Carlo error over time [Gordon, Salmond and Smith (1993)]. In particular, without resampling, the largest normalised weight of the particle sample converges quickly to one as the number of iterations increases [Del Moral and Doucet (2003)]. This means that most of the computational effort is wasted on particles that contribute little to the end results.

Resampling also appears in survey sampling under the name of “unequal probability sampling” [Tillé (2006)], but in a context slightly different from the one we consider in this paper. In survey sampling, only  $M < N$  “units” are selected and the object of interest after the (re)sampling operation, the Horvitz–Thompson empirical process [HTEP, see, e.g., Bertail, Chautru and Cléménçon (2017)] is another un-normalized weighted sum of Dirac measures. Adapting the statement and the assumptions of our first main result, Theorem 1 in Section 2, in order to study the asymptotic behaviour of the HTEP is possible but beyond the scope of this paper. Yet another statistical procedure where resampling appears is the the weighted bootstrap [Barbe and Bertail (1995)].

There are various existing resampling methods. Multinomial resampling is perhaps the simplest technique, where given the weights, the indices  $A^n$  are generated conditionally independently from the finite distribution that assigns probability  $W^n$  to outcome  $n$ . In particle filtering it is common practice to replace multinomial resampling with techniques which are computationally faster and empirically more accurate. However, these advanced resampling techniques are generally not straightforward to analyse because they induce complicated dependence between output samples, and various aspects of their behaviour are still not understood.

Following definitions and an account of what is known about existing resampling techniques, our first main result, Theorem 1 in Section 2, is a general consistency result for resampling based on the notion of *negative association* [Joag-Dev and Proschan (1983)]. An application of this theorem gives, to our knowledge, the first proof of almost sure weak convergence of the random probability measures output from the stratified resampling method of Kitagawa (1996). A notable feature of Theorem 1 is that, although its assumptions do not require the input particles to be algorithmically ordered in a particular way, its proof involves establishing a necessary and sufficient condition for almost sure weak convergence involving ordering using the Hilbert space-filling curve. Here, we build on Gerber and Chopin (2015), who used the Hilbert curve to derive and analyse a quasi-Monte Carlo version of sequential Monte Carlo samplers.

The systematic resampling method of Carpenter, Clifford and Fearnhead (1999), which involves a sampling technique first proposed by Madow and Madow (1944), is a very popular and computationally cheap resampling technique, with the property that the number of offspring of any sample with weight  $W$  in a population of size  $N$  is with probability 1 either  $\lfloor NW \rfloor$  or  $\lfloor NW \rfloor + 1$ . However,

depending on the order of the input particles, the error variance for systematic resampling can fail to converge to zero as  $N \rightarrow +\infty$ ; see Douc, Cappé and Moulines (2005) and L’Ecuyer and Lemieux (2000). We complement this insight by providing a counter-example to almost sure weak convergence. We then introduce a new resampling method, called Srinivasan Sampling Process (SSP) resampling, which corrects this deficiency: it also has the property that offspring numbers are of the form either  $\lfloor NW \rfloor$  or  $\lfloor NW \rfloor + 1$ , but it provably converges irrespective of the order of input particles, by another application of our Theorem 1.

Kitagawa (1996) conjectured that in the case that the state-space is  $\mathbb{R}$ , ordering the particles input to stratified resampling according to their states leads to faster convergence. In particular, he suggested that the integrated square error between empirical cdf’s before and after resampling behaves as  $\mathcal{O}(N^{-2})$ , compared to the standard Monte Carlo rate  $\mathcal{O}(N^{-1})$  in the un-ordered case. We confirm this conjecture by proving, under mild conditions, that for stratified resampling on state-space  $\mathbb{R}^d$  with input particles ordered by their states using the Hilbert curve, the variance of the resampling error is  $\mathcal{O}(N^{-(1+1/d)})$ . Kitagawa also examined the behaviour of a deterministic resampling scheme; we identify the variant of it which is optimal in terms of the Kolmogorov metric when the state-space is  $\mathbb{R}$ . We also prove the almost sure weak consistency of stratified and systematic when the particles are Hilbert-ordered.

Finally, we discuss the implications of our results on particle filtering. In particular, we show that particle estimates are consistent when resampling schemes such as, for example, SSP or stratified resampling are used. In addition, we show that the ordered version of stratified resampling dominates other resampling schemes in terms of asymptotic variance of particle estimates.

All the proofs are gathered in the Supplementary Material [Gerber, Chopin and Whiteley (2019)].

**2. Preliminaries.**

2.1. *Notation and conventions.* Let  $\mathcal{X}$  be an open subset of  $\mathbb{R}^d$ ,  $\mathbb{X}$  its Borel  $\sigma$ -algebra,  $\mathcal{P}(\mathcal{X})$  the set of probability measures on  $(\mathcal{X}, \mathbb{X})$ ,  $\mathcal{P}_b(\mathcal{X}) \subset \mathcal{P}(\mathcal{X})$  the subset of measures in  $\mathcal{P}(\mathcal{X})$  which admit a continuous and bounded density with respect to  $\lambda_d$ , the Lebesgue measure on  $\mathcal{X}$  and  $\mathcal{P}_f(\mathcal{X}) \subset \mathcal{P}(\mathcal{X})$  the subset of measures in  $\mathcal{P}(\mathcal{X})$  whose support is a finite set.

For integers  $1 \leq a \leq b$ , we will often use the index shorthand  $z_{a:b} = (z_a, \dots, z_b)$  and  $z^{a:b} = (z^a, \dots, z^b)$ , and let  $1 : N = \{1, \dots, N\}$ .

For any measurable mapping  $\varphi$  from  $(\mathcal{X}, \mathbb{X})$  to some measurable space  $(\mathcal{Y}, \mathbb{Y})$  and a probability measure  $\pi \in \mathcal{P}(\mathcal{X})$ , we write  $\pi_\varphi$  for the push-forward of  $\pi$  by  $\varphi$ . The set of continuous and bounded functions on  $\mathcal{X}$  is denoted by  $\mathcal{C}_b(\mathcal{X})$  and we use the symbol “ $\xrightarrow{w}$ ” to denote weak convergence; that is, for sequence  $(\pi^N)_{N \geq 1}$  in  $\mathcal{P}(\mathcal{X})$  and  $\pi \in \mathcal{P}(\mathcal{X})$ ,

$$\pi^N \xrightarrow{w} \pi \iff \lim_{N \rightarrow +\infty} \int \varphi d\pi^N = \int \varphi d\pi, \quad \forall \varphi \in \mathcal{C}_b(\mathcal{X}).$$

Throughout the paper, we consider a fixed probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  on which all random variables are defined. With  $\mathcal{B}([0, 1]^{\mathbb{N}})$  denoting the Borel  $\sigma$ -algebra on  $[0, 1]^{\mathbb{N}}$ , let  $U = (U_1, U_2, \dots)$  be a  $([0, 1]^{\mathbb{N}}, \mathcal{B}([0, 1]^{\mathbb{N}}))$ -valued random variable on  $(\Omega, \mathcal{F}, \mathbb{P})$ , such that  $\mathbb{P}$  makes  $(U_1, U_2, \dots)$  independent of each other and all other random variables, and such that each  $U_i$  is distributed uniformly on  $[0, 1]$ .

We note that one can choose a countable subset of  $\mathcal{C}_b(\mathcal{X})$  that completely determines weak convergence, hence for random measures  $(\pi^N)_{N \geq 1}$ , the event  $\{\pi^N \xrightarrow{w} \pi\}$  is measurable.

For  $\pi \in \mathcal{P}(\mathcal{X})$ , we denote by  $\pi(\varphi)$  the expectation  $\int_{\mathcal{X}} \varphi(x)\pi(dx)$ , and for a random variable  $Z = (Z_1, \dots, Z_d)$  whose distribution is  $\pi$  we denote by  $F_{\pi}(a) = \mathbb{P}(Z_1 \leq a_1, \dots, Z_d \leq a_d)$ ,  $a = (a_1, \dots, a_d)$ , its CDF (cumulative distribution function) and, when  $d = 1$ , by  $F_{\pi}^{-}$  its generalized inverse:  $F_{\pi}^{-}(u) = \inf\{x : F_{\pi}(x) \geq u\}$ .

For each  $N \geq 1$ , we consider a distinguished collection of random variables  $\zeta^N = (X^{n,N}, W^{n,N})_{n=1}^N$ , with each  $X^{n,N}$  valued in  $\mathcal{X}$ , each  $(W^{n,N})_{n=1}^N$  valued in  $\mathbb{R}_+$ , and such that  $\mathbb{P}$ -a.s.,  $\sum_{n=1}^N W^{n,N} = 1$ . When no confusion may arise, we suppress dependence on  $N$  and write  $\zeta^N = (X^n, W^n)_{n=1}^N$ . We associate with  $\zeta^N$  the random measure  $\pi^N = \sum_{n=1}^N W^n \delta(X^n)$ , the (random) CDF

$$F_N(n) = \sum_{m=1}^N W^m \mathbb{1}(m \leq n), \quad n \in 1 : N,$$

and its inverse is denoted  $F_N^{-}$ .

To lighten notation, we shall write  $\mathbb{P}_{\zeta^N}(\cdot)$ ,  $\mathbb{E}_{\zeta^N}[\cdot]$ ,  $\text{Var}_{\zeta^N}[\cdot]$ ,  $\text{Cov}_{\zeta^N}[\cdot, \cdot]$  for conditional probability, expectation, variance and covariance given  $\zeta^N$ .

Let  $\mathcal{Z}^N = \{(x, w) \in \mathcal{X}^N \times \mathbb{R}_+^N : \sum_{n=1}^N w_n = 1\}$  and define the disjoint union  $\mathcal{Z} := \bigcup_{N=1}^{+\infty} \mathcal{Z}^N$ . So we may think of  $\zeta^N$  as a random point in  $\mathcal{Z}^N$ , and hence  $\mathcal{Z}$ .

**DEFINITION 1.**  $\mathcal{X} \subseteq \mathbb{R}^d$  is said to be cubifiable if there exist measurable sets  $\mathcal{X}_i \subseteq \mathbb{R}$ ,  $i = 1, \dots, d$ , such that:

1.  $\mathcal{X} = \times_{i=1}^d \mathcal{X}_i$ ;
2. for any  $i \in 1 : d$ , there exists a  $C^1$ -diffeomorphism  $\psi_i : \mathcal{X}_i \rightarrow (0, 1)$  which is strictly increasing on  $\mathcal{X}_i$ .

We shall write  $\psi(x) = (\psi_1(x_1), \dots, \psi_d(x_d))$ ,  $x = x_{1:d} \in \mathcal{X}$ , the resulting  $C^1$ -diffeomorphism from  $\mathcal{X}$  into  $(0, 1)^d$ .

We recall for the reader that function  $\psi : \mathcal{X} \rightarrow (0, 1)^d$  is a  $C^1$ -diffeomorphism if it is a bijection and its inverse  $\psi^{-1} : (0, 1)^d \rightarrow \mathcal{X}$  is continuously differentiable. In what follows, for a cubifiable set  $\mathcal{X}$  we denote by  $\mathcal{D}(\mathcal{X})$  the set of all  $C^1$ -diffeomorphisms from  $\mathcal{X}$  into  $(0, 1)^d$  that verify the conditions of Definition 1.

Cubifiable sets are sets that can be written as  $\mathcal{X} = \times_{i=1}^d (a_i, b_i)$  for some  $a_i, b_i \in \mathbb{R} \cup \{-\infty, +\infty\}$ . The point of these sets is to be able to work “as if”  $\mathcal{X} = (0, 1)^d$ .

The hypercube  $(0, 1)^d$  will play a key role below because the Hilbert space-filling curve, which is essential in this work, is defined on this hypercube.

Most of the results presented below assume that the limiting distribution  $\pi$  admits a continuous and bounded density. Consequently, to work “as if”  $\mathcal{X} = (0, 1)^d$  we will often assume that  $\pi$  belongs to

$$\tilde{\mathcal{P}}_b(\mathcal{X}) = \{\pi \in \mathcal{P}_b(\mathcal{X}) : \exists \psi \in \mathcal{D}(\mathcal{X}) \text{ s.t. } \pi_\psi \in \mathcal{P}_b((0, 1)^d)\}.$$

The following result provides a sufficient condition to have  $\pi \in \tilde{\mathcal{P}}_b(\mathcal{X})$ . We denote by  $p_\pi$  the density (w.r.t.  $\lambda_d$ ) of  $\pi \in \mathcal{P}_b(\mathcal{X})$  and, for  $I \subset 1 : d$ , we write  $x_I = (x_i, i \in I)$  and  $x_{\setminus I} = (x_i, i \notin I)$ .

**LEMMA 1.** *Let  $\mathcal{X}$  be a cubifiable set,  $\delta > 0$  and  $\pi \in \mathcal{P}_b(\mathcal{X})$  such that  $\forall I \subseteq 1 : d$  and  $\forall x_{\setminus I} \in \times_{i \notin I} \mathcal{X}_i$  we have  $\sup_{x_I \in \times_{i \in I} \mathcal{X}_i} p_\pi(x) \prod_{i \in I} |x_i|^{1+\delta} \leq C$  for some  $C < +\infty$ . Then  $\pi \in \tilde{\mathcal{P}}_b(\mathcal{X})$ .*

Recall that  $\sup_{x \in \mathbb{R}} p_\pi(x)|x| < +\infty$  for any  $\pi \in \mathcal{P}_b(\mathbb{R})$ . Therefore, as  $\delta > 0$  is arbitrary in the lemma, very few extra conditions on the tails of  $\pi \in \mathcal{P}_b(\mathbb{R})$  are needed in order to have  $\pi \in \tilde{\mathcal{P}}_b(\mathbb{R})$  when  $d = 1$ . When  $d > 1$ , assuming that  $\pi \in \tilde{\mathcal{P}}_b(\mathbb{R}^d)$  is more restrictive since the lemma requires some uniformity in the behaviour of tails. However, we note that members of  $\tilde{\mathcal{P}}_b(\mathbb{R}^d)$  may not have a first moment and, therefore, the sufficient condition of Lemma 1 appears to be quite weak.

2.2. Resampling schemes: Definitions and properties.

**DEFINITION 2.** A resampling scheme is a mapping  $\rho : [0, 1]^{\mathbb{N}} \times \mathcal{Z} \rightarrow \mathcal{P}_f(\mathcal{X})$  such that, for any  $N \geq 1$  and  $z = (x^n, w^n)_{n=1}^N \in \mathcal{Z}^N$ ,

$$\rho(u, z) = \frac{1}{N} \sum_{n=1}^N \delta(x^{a_N^n(u, z)}),$$

where for each  $n$ ,  $a_N^n : [0, 1]^{\mathbb{N}} \times \mathcal{Z}^N \rightarrow 1 : N$  is a measurable function.

Given  $u \in [0, 1]^{\mathbb{N}}$ , the mapping  $\rho(u, \cdot)$  therefore takes as input a weighted point set  $z = (x^n, w^n)_{n=1}^N$ , selects  $N$  indices  $(a_N^n(u, z))_{n=1}^N$  in the set  $1 : N$  and returns a probability measure on  $(x^n)_{n=1}^N$  with the property that each  $x^{a_N^n(u, z)}$  has weight  $N^{-1}$ .

Instances of the function  $a_N^n$  are given below. We shall use the shorthand  $\rho(z)$  for the random measure  $\rho(U, z)$ ,  $z \in \mathcal{Z}^N$ , and  $A^n$  for the random indices  $a_N^n(U, \zeta^N)$ . Introducing the quantities

$$(2) \quad \#^n(u, z) = \text{card}\{i \in 1 : N \text{ s.t. } a_N^i(u, z) = n\}, \quad \Delta_{\rho, z}^n = \#^n(U, z) - Nw^n,$$

a resampling scheme  $\rho$  is said to be *unbiased* if, for any  $N \geq 1$ ,  $n \in 1 : N$  and  $z \in \mathcal{Z}^N$ ,

$$\mathbb{E}[\Delta_{\rho,z}^n] = 0.$$

We now define the resampling schemes of primary interest in this work.

- *Multinomial resampling*:  $\rho_{\text{multi}}$  such that

$$a_N^n(u, \zeta^N) = F_N^-(u_n).$$

In this case, the  $a_N^n(U, \zeta^N)$  are i.i.d. (independent and identically distributed) draws from the distribution which assigns probability  $W^n$  to outcome  $n$ .

- *Stratified resampling*:  $\rho_{\text{strat}}$  such that

$$a_N^n(u, \zeta^N) = F_N^-\left(\frac{n-1+u_n}{N}\right).$$

- *Systematic resampling*:  $\rho_{\text{syst}}$  such that

$$a_N^n(u, \zeta^N) = F_N^-\left(\frac{n-1+u_1}{N}\right).$$

The following definition captures the notion of almost sure weak convergence of the random measures  $(\pi^N)_{N \geq 1}$  which we shall study and is similar to condition (9) in [Crisan and Doucet (2002)].

DEFINITION 3. Let  $\mathcal{P}_0 \subseteq \mathcal{P}(\mathcal{X})$ . Then we say that a resampling scheme  $\rho : [0, 1]^N \times \mathcal{Z} \rightarrow \mathcal{P}_f(\mathcal{X})$  is  $\mathcal{P}_0$ -consistent if, for any  $\pi \in \mathcal{P}_0$  and  $(\zeta^N)_{N \geq 1}$  such that  $\pi^N \xrightarrow{w} \pi$ ,  $\mathbb{P}$ -a.s., one has

$$\rho(\zeta^N) \xrightarrow{w} \pi, \quad \mathbb{P}\text{-a.s.}$$

It is well known that multinomial, stratified and systematic resampling are unbiased. An account of various properties of these methods can be found in Douc, Cappé and Moulines (2005).

Crisan and Doucet [(2002), Lemma 2] shows that multinomial resampling is  $\mathcal{P}(\mathcal{X})$ -consistent for any measurable set  $\mathcal{X} \subseteq \mathbb{R}^d$ .

It is easy to show [Stein (1987), Douc, Cappé and Moulines (2005)] that stratified resampling dominates multinomial resampling in terms of variance, that is,

$$\text{Var}[\rho_{\text{strat}}(z)(\varphi)] \leq \text{Var}[\rho_{\text{multi}}(z)(\varphi)], \quad \forall z \in \mathcal{Z}$$

for any measurable  $\varphi : \mathcal{X} \rightarrow \mathbb{R}$ . Similar results are harder to derive for systematic resampling, owing to the strong dependencies between the resampled indices. However, it is known [Douc, Cappé and Moulines (2005)] that the variance of  $\rho_{\text{syst}}(\zeta^N)(\varphi)$  may not converge to 0 as  $N \rightarrow +\infty$  [see also L'Ecuyer and Lemieux (2000) for an explanation of this phenomenon].

### 3. Convergence of resampling schemes based on negative association.

3.1. *A general consistency result.* Before stating the main result of this section we recall the definition of negatively associated (NA) random variables [Joag-Dev and Proschan (1983)].

DEFINITION 4. A collection of random variables  $(Z^n)_{n=1}^N$  are negatively associated if, for every pair of disjoint subsets  $I_1$  and  $I_2$  of  $\{1, \dots, N\}$ ,

$$\text{Cov}(\varphi_1(Z^n, n \in I_1), \varphi_2(Z^n, n \in I_2)) \leq 0$$

for all coordinatewise nondecreasing functions  $\varphi_1$  and  $\varphi_2$  such that for  $k \in \{1, 2\}$ ,  $\varphi_k : \mathbb{R}^{|I_k|} \rightarrow \mathbb{R}$ , and such that the covariance is well defined.

THEOREM 1. Let  $\mathcal{X}$  be a cubifiable set and  $\rho$  be an unbiased resampling scheme such that the following conditions hold:

(H<sub>1</sub>) For any  $N \geq 1$  and  $z \in \mathcal{Z}^N$ , the random variables  $(\#^n(U, z))_{n=1}^N$  are negatively associated.

(H<sub>2</sub>) There exists a sequence  $(r_N)_{N \geq 1}$  of nonnegative real numbers such that  $r_N = o(N / \log N)$  and, for  $N$  large enough,

$$\sup_{z \in \mathcal{Z}^N} \sum_{n=1}^N \mathbb{E}[(\Delta_{\rho, z}^n)^2] \leq r_N N, \quad \sum_{N=1}^{\infty} \sup_{z \in \mathcal{Z}^N} \mathbb{P}\left(\max_{n \in 1:N} |\Delta_{\rho, z}^n| > r_N\right) < +\infty.$$

Then  $\rho$  is  $\tilde{\mathcal{P}}_b(\mathcal{X})$ -consistent.

The strategy of the proof is the following. In a first step, we show that when  $\sigma_N^*$  is a permutation of  $1 : N$  which corresponds to ordering input particles using the Hilbert space filling curve (details of which we postpone to Section 4), the resampling scheme  $\rho$  is  $\tilde{\mathcal{P}}_b(\mathcal{X})$ -consistent if and only if

$$(3) \quad \lim_{N \rightarrow +\infty} \max_{m \in 1:N} \left| \sum_{n=1}^m \Delta_{\rho, z_N}^{\sigma_N^*(n)} \right| = 0, \quad \mathbb{P}\text{-a.s.}$$

for any sequence  $(z_N)_{N \geq 1}$  with  $z_N \in \mathcal{Z}^N$ . In a second step, we show that the hypotheses (H<sub>1</sub>) and (H<sub>2</sub>) are sufficient to establish (3), via a maximal inequality for negatively associated random variables due to Shao (2000). We stress here that the permutation  $\sigma_N^*$  is introduced solely as a device in the proof; there is no assumption in Theorem 1 that the input particles are algorithmically sorted in any particular way. The reader should note, in fact, that (H<sub>1</sub>) must hold for all  $z$ , and (H<sub>2</sub>) is uniform in  $z$ , and hence all permutations of the input particles.

3.2. *Discussion of (H<sub>1</sub>) and (H<sub>2</sub>).* From the definition of  $\#^n(U, z)$  given in (2), it follows that  $\sum_{n=1}^N \#^n(U, z) = N$ ,  $\mathbb{P}$ -a.s. Intuitively, this constraint suggests that at least some random variables in the set  $(\#^n(U, z))_{n=1}^N$  are negatively correlated. (H<sub>1</sub>) may be understood as imposing that *all* these random variables are negatively correlated.

(H<sub>2</sub>) alone is not sufficient to guarantee the consistency of an unbiased resampling scheme. If a resampling scheme  $\rho$  violates (H<sub>1</sub>), then it is indeed possible to find examples where the offspring numbers are positively correlated in a way that, with positive probability, prevents the limit in (3) from being zero. The next result formalizes this assertion in the context of systematic resampling. Its proof involves a somewhat technical construction of a counter-example.

PROPOSITION 1. *The systematic resampling scheme  $\rho_{\text{sys}}$  is unbiased, satisfies (H<sub>2</sub>) with  $r_N = 1$  but is not  $\tilde{\mathcal{P}}_b(\mathcal{X})$ -consistent.*

On the other hand, (H<sub>1</sub>) alone is not enough to guarantee consistency. If we consider the resampling scheme  $\rho$  such that  $\#^n(U, z) = N$  with probability  $w^n$ , it is easily checked  $\rho$  is unbiased and (H<sub>1</sub>) holds, but this resampling scheme is obviously not  $\tilde{\mathcal{P}}_b(\mathcal{X})$ -consistent. (H<sub>2</sub>) rules out this kind of situation via constraints on the second moments and negligibility of the deviations of the offspring numbers  $(\#^n(U, z))_{n=1}^N$  from their respective means  $(Nw^n)_{n=1}^N$ .

3.3. *Some comments about systematic resampling.* Systematic resampling has the property that  $\#^n(U, z)$  is either  $\lfloor Nw^n \rfloor$  or  $\lfloor Nw^n \rfloor + 1$ ,  $\mathbb{P}$ -a.s., hence  $|\Delta^n_{\rho_{\text{sys}}, z^N}| \leq 1$ ,  $\mathbb{P}$ -a.s., so that (H<sub>2</sub>) holds with  $r_N = 1$  as stated in Proposition 1.

A corollary of this latter is that systematic resampling violates (H<sub>1</sub>). A simple way to establish this result is to take a  $z \in \mathcal{Z}^N$  such that we have  $Nw^n - \lfloor Nw^n \rfloor = 1/2$  for  $n = 1, \dots, 3$ . Then

$$\mathbb{P}(\#^1(U, z) = \#^3(U, z) = 1) = \frac{1}{2} > \mathbb{P}(\#^1(U, z) = 1)\mathbb{P}(\#^3(U, z) = 1) = \frac{1}{4}$$

showing that the collection of random variables  $(\#^n(U, z))_{n=1}^N$  is not NA.

To overcome the lack of consistency (in the sense of Definition 3) of systematic resampling, we introduce below (Section 3.4.3) a new resampling scheme, named SSP (for Srinivasan Sampling Process) resampling, which both satisfies the NA condition (H<sub>1</sub>) and shares the property of systematic resampling that  $|\Delta^n_{\rho_{\text{sys}}, z^N}| \leq 1$  for all  $n \in 1 : N$ ,  $\mathbb{P}$ -a.s., so that (H<sub>2</sub>) also holds with  $r_N = 1$  for this new resampling scheme.

### 3.4. Applications of Theorem 1.

3.4.1. *Multinomial resampling.* As already mentioned, it is a known result that multinomial resampling is  $\mathcal{P}(\mathcal{X})$ -consistent for any measurable  $\mathcal{X} \subseteq \mathbb{R}^d$  [Crisan



and Doucet (2002), Lemma 2]. Theorem 1 may be applied to obtain a similar result.

**COROLLARY 1.** *Let  $\mathcal{X}$  be a cubifiable set. Then the resampling scheme  $\rho_{\text{multi}}$  verifies conditions  $(H_1)$  and  $(H_2)$  of Theorem 1 and is therefore  $\tilde{\mathcal{P}}_b(\mathcal{X})$ -consistent.*

Condition  $(H_1)$  holds for multinomial resampling as shown by Joag-Dev and Proschan (1983) while  $(H_2)$  is verified using properties of the binomial distribution and Hoeffding’s inequality.

For similar reasons, the conditions of Theorem 1 are also satisfied by the residual resampling scheme of Liu and Chen (1998).

**3.4.2. Stratified resampling.** To the best of our knowledge, the following corollary of Theorem 1 is the first almost sure weak convergence result for Kitagawa’s (1996) stratified resampling scheme.

**COROLLARY 2.** *Let  $\mathcal{X}$  be a cubifiable set. Then the resampling scheme  $\rho_{\text{strat}}$  verifies conditions  $(H_1)$  and  $(H_2)$  of Theorem 1 and is therefore  $\tilde{\mathcal{P}}_b(\mathcal{X})$ -consistent.*

Verifying  $(H_1)$  in this situation involves the observation that stratified resampling is a “Balls and Bins” experiment [Dubhashi and Ranjan (1998)] in which  $N$  balls are independently thrown into  $N$  bins, the total number of balls occupying the  $n$ th bin is  $\#^n(U, z)$ , and where the probability of falling in a given bin varies across balls, due to the stratified nature of the sampling. The fact that  $(H_1)$  holds is then a direct consequence of Theorem 14 in Dubhashi and Ranjan (1998), which establishes the NA of occupancy numbers in a slightly more general balls and bins problem where the number of balls is not necessarily equal to the number of bins.  $(H_2)$  holds because  $|\Delta_{\rho_{\text{strat}}, z}| \leq 2$ ,  $\mathbb{P}$ -a.s.

It is worth noting that the conditions of Theorem 1 are also satisfied by the stratified version of the residual resampling scheme of Liu and Chen (1998), where the multinomial resampling part is replaced by a stratified resampling step. Denoting these two resampling schemes by  $\rho_{\text{res/multi}}$  and  $\rho_{\text{res/strat}}$ , respectively, the stratified version of residual resampling has the interesting property that, for any measurable  $\varphi : \mathcal{X} \rightarrow \mathbb{R}$  we have [see Douc, Cappé and Moulines (2005) for the second inequality]

$$\text{Var}[\rho_{\text{res/strat}}(z)(\varphi)] \leq \text{Var}[\rho_{\text{res/multi}}(z)(\varphi)] \leq \text{Var}[\rho_{\text{multi}}(z)(\varphi)], \quad \forall z \in \mathcal{Z}.$$

In addition,  $\rho_{\text{res/strat}}$  has the advantage to be easier and slightly cheaper to implement than  $\rho_{\text{res/multi}}$ .

**3.4.3. SSP resampling.** The underlying idea of SSP resampling is to see the resampling scheme as a rounding operation, where the vector of “weights”  $(Nw^1, \dots, Nw^N)$  is  $\mathbb{P}$ -a.s. transformed into a point  $(Y_1(U), \dots, Y_N(U))$  in  $\mathbb{N}^N$  satisfying the constraint  $\sum_{n=1}^N Y_n(U) = N$ .

Before proceeding further, we recall the terminology that, for  $\xi \in \mathbb{R}_+$ , a random variable  $Y : \Omega \rightarrow \mathbb{N}$  is called a randomized rounding of  $\xi$  if

$$\mathbb{P}(Y = \lfloor \xi \rfloor + 1) = \xi - \lfloor \xi \rfloor, \quad \mathbb{P}(Y = \lfloor \xi \rfloor) = 1 - (\xi - \lfloor \xi \rfloor).$$

Hence, any algorithmic technique for constructing randomized roundings that takes as input a vector  $(\xi_1, \dots, \xi_N) \in \mathbb{R}_+^N$  and returns  $\mathbb{P}$ -a.s. as output a vector  $(Y_1(U), \dots, Y_N(U)) \in \mathbb{N}^N$  verifying  $\sum_{n=1}^N Y_n(U) = \sum_{n=1}^N \xi_n$  may be used to construct an unbiased resampling mechanism; systematic resampling can be viewed as being constructed in this way.

The SSP resampling scheme  $\rho_{\text{ssp}} : [0, 1]^{\mathbb{N}} \times \mathcal{Z} \rightarrow \mathcal{P}_f(\mathcal{X})$  is based on the Srinivasan’s (2001) randomized rounding technique [also known as pivotal sampling in the sampling survey literature, see, e.g., Deville and Tillé (1998)] and is presented in Algorithm 1. To see that this latter indeed defines a randomized rounding process it suffices to note that step (2) leaves unchanged the expectation of the vector  $(Y_{\text{ssp}}^n(U))_{n=1}^N$  while, by construction, each iteration of the algorithm leaves the quantity  $\sum_{n=1}^N Y_{\text{ssp}}^n(U)$  unchanged with  $\mathbb{P}$ -probability one. By Dubhashi, Jonasson and Ranjan [(2007), Theorem 5.1; see also Brown Kramer, Cutler and Radcliffe (2011)], the collection of random variables  $(Y_{\text{ssp}}^n(U))_{n=1}^N$  produced by the SSP described in Algorithm 1 is NA. Together with Theorem 1, this result allows to readily show the consistency of  $\rho_{\text{ssp}}$ .

**Algorithm 1** SSP resampling

**Inputs:**  $u \in [0, 1]^{\mathbb{N}}$  and  $(\xi_1, \dots, \xi_N) \in \mathbb{R}_+^N$  such that  $\sum_{n=1}^N \xi_n \in \mathbb{N}$ .

**Output:**  $(Y_{\text{ssp}}^1(u), \dots, Y_{\text{ssp}}^N(u)) \in \mathbb{N}^N$  such that  $\sum_{n=1}^N Y_{\text{ssp}}^n(u) = \sum_{n=1}^N \xi_n$ .

Initialization:  $(Y_{\text{ssp}}^1(u), \dots, Y_{\text{ssp}}^N(u)) \leftarrow (\xi_1, \dots, \xi_N)$ ,  $(n, m, k) \leftarrow (1, 2, 1)$

Iterate the following steps until  $(Y_{\text{ssp}}^1(u), \dots, Y_{\text{ssp}}^N(u)) \in \mathbb{N}^N$ :

- (1) Let  $\delta$  be the smallest number in  $(0, 1)$  such that at least one of  $Y_{\text{ssp}}^n(u) + \delta$  or  $Y_{\text{ssp}}^m(u) - \delta$  is an integer, and let  $\epsilon$  be the smallest number in  $(0, 1)$  such that at least one of  $Y_{\text{ssp}}^n(u) - \epsilon$  or  $Y_{\text{ssp}}^m(u) + \epsilon$  is an integer.
- (2) If  $u_k \leq \epsilon / (\delta + \epsilon)$  set  $(Y_{\text{ssp}}^n(u), Y_{\text{ssp}}^m(u)) \leftarrow (Y_{\text{ssp}}^n(u) + \delta, Y_{\text{ssp}}^m(u) - \delta)$ ; otherwise set  $(Y_{\text{ssp}}^n(u), Y_{\text{ssp}}^m(u)) \leftarrow (Y_{\text{ssp}}^n(u) - \epsilon, Y_{\text{ssp}}^m(u) + \epsilon)$ .
- (3) Update  $n$  and  $m$  as follows:
  1. If  $(Y_{\text{ssp}}^n(u), Y_{\text{ssp}}^m(u)) \in \mathbb{N}^2$ ,  $(n, m) \leftarrow (m + 1, m + 2)$ ;
  2. If  $Y_{\text{ssp}}^n(u) \in \mathbb{N}$  and  $Y_{\text{ssp}}^m(u) \notin \mathbb{N}$  set  $(n, m) \leftarrow (m, m + 1)$ ;
  3. If  $Y_{\text{ssp}}^n(u) \notin \mathbb{N}$  and  $Y_{\text{ssp}}^m(u) \in \mathbb{N}$  set  $(n, m) \leftarrow (m, m + 1)$ .
- (4)  $k \leftarrow k + 1$

**COROLLARY 3.** *Let  $\mathcal{X}$  be a cubifiable set. Then the resampling scheme  $\rho_{\text{SSP}}$  verifies conditions  $(\text{H}_1)$  and  $(\text{H}_2)$  of Theorem 1 and is therefore  $\tilde{\mathcal{P}}_b(\mathcal{X})$ -consistent.*

Algorithm 1 has complexity  $\mathcal{O}(N)$ , like other standard resampling schemes. An open question is whether or not SSP resampling dominates multinomial resampling in terms of variance. See Section 5.5 for a numerical comparison.

Lastly in this section, we note that a resampling scheme proposed in Crisan (2001) may also be interpreted as a randomized rounding technique. However, to the best of our knowledge, there are no convergence results for this resampling scheme.

**4. Convergence of ordered resampling schemes.** Kitagawa [(1996), Appendix A] provided numerical results about the behaviour of stratified resampling in the case that  $d = 1$  and the input particles are ordered according to their states. He conjectured that in this situation, the error of stratified resampling is of size  $\mathcal{O}(N^{-2})$ , compared to  $\mathcal{O}(N^{-1})$  without the ordering. He also considered a deterministic resampling scheme, and found that in same  $d = 1$  case and with ordered particles, it also exhibited  $\mathcal{O}(N^{-2})$  convergence.

The purpose of this section is to provide a rigorous investigation of this topic. While Kitagawa (1996) measured the error introduced by a resampling scheme by the integrated square error between empirical CDF’s before and after resampling, we compare below the probability measures before and after resampling by comparing their expectations for some test functions. Notably, we present in this section results on the convergence rate of the variance of stratified resampling when applied on ordered input particles. We first consider the case  $d = 1$  and then the general  $d \geq 1$  case in which particles input to resampling are ordered using the Hilbert space filling curve.

**4.1. Ordered resampling schemes on univariate sets.** In this subsection, we present results for a univariate set  $\mathcal{X}$ , which is the setup considered by Kitagawa (1996). The existence of a natural order in this context greatly facilitates the presentation and allows to derive more precise convergence results than in multivariate settings.

We denote below by  $\rho_{\text{strat}}^*$  the ordered stratified resampling scheme; that is,  $\rho_{\text{strat}}^* : [0, 1]^{\mathbb{N}} \times \mathcal{Z} \rightarrow \mathcal{P}_f(\mathcal{X})$  is defined by

$$\rho_{\text{strat}}^*(u, z) = \rho_{\text{strat}}(u, (z_{\sigma_N^*(n)})_{n=1}^N), \quad (u, z) \in [0, 1]^{\mathbb{N}} \times \mathcal{Z}^N$$

with  $\sigma_N^*$  a permutation of  $1 : N$  such that  $z_{\sigma_N^*(1)} \leq \dots \leq z_{\sigma_N^*(N)}$ . In words,  $\rho_{\text{strat}}^*(\zeta^N)$  simply amounts to apply the stratified resampling scheme  $\rho_{\text{strat}}$  on the ordered input point set  $\zeta^{N, \sigma_N^*} := (X^{\sigma_N^*(n)}, W^{\sigma_N^*(n)})_{n=1}^N$ . Notice that  $\rho_{\text{strat}}^*(\zeta^N)$  is such that

$$(4) \quad X^{A^n} = F_{\pi^N}^- \left( \frac{n - 1 + U_n}{N} \right), \quad n \in 1 : N;$$

that is, the resampled particles  $(X^{A^n})_{n=1}^N$  are obtained by sampling from the empirical distribution  $\pi^N$  using the stratified point set  $((n - 1 + U_n)/N)_{n=1}^N$ .

The following theorem shows that under mild conditions the variance induced by ordered stratified resampling converges faster than  $N^{-1}$ . In addition, it also provides conditions under which one has a nonasymptotic bound of size  $N^{-2}$  for this resampling method.

**THEOREM 2.** *Let  $\mathcal{X} \subseteq \mathbb{R}$  be a cubifiable set. Then the following results hold:*

1. *Let  $\pi \in \tilde{\mathcal{P}}_b(\mathcal{X})$  have a strictly positive density and  $(\zeta^N)_{N \geq 1}$  be such that  $\pi^N \xrightarrow{w} \pi$ ,  $\mathbb{P}$ -a.s., and such that,  $\lim_{N \rightarrow +\infty} (\max_{n \in 1:N} W^{n,N}) = 0$ ,  $\mathbb{P}$ -a.s. Then, for any  $\varphi \in \mathcal{C}_b(\mathcal{X})$ ,  $\text{Var}_{\zeta^N}[\rho_{\text{strat}}^*(\zeta^N)(\varphi)] = o(1/N)$ ,  $\mathbb{P}$ -a.s.*

2. *Let  $\varphi : \mathcal{X} \rightarrow \mathbb{R}$  be a continuously differentiable function such that, for a  $\delta > 0$ , we have  $\sup_{x \in \mathcal{X}} \frac{d\varphi}{dx}(x)|x|^{1+\delta} < +\infty$ . Then there exists a constant  $C_\varphi < +\infty$  such that, for all  $N \geq 1$ ,*

$$\text{Var}[\rho_{\text{strat}}^*(z)(\varphi)] \leq C_\varphi N^{-2}, \quad \forall z \in \mathcal{Z}^N.$$

The second observation of Kitagawa [(1996), page 23] is that deterministic resampling mechanisms may be used when applied to the ordered input particles  $\zeta^{N, \sigma_N^*}$ . In particular, he considered a resampling scheme defined by (4) but with the random variables  $(U_n)_{n=1}^N$  replaced by a deterministic point in  $\alpha \in (0, 1)$ . In the notation of this work, for  $\alpha \in (0, 1)$  Kitagawa (1996) considered the resampling scheme  $\rho_\alpha^*$  defined by  $\rho_\alpha^*(u, z) = \rho_{\text{strat}}^*(\alpha_\infty, z)$  with  $\alpha_\infty$  the vector in  $(0, 1)^N$  having  $\alpha$  in all its entries. The consistency of this deterministic resampling mechanism trivially follows from Corollary 4 (see below) and the fact that [Niederreiter (1992), Theorem 2.6, page 15]

$$(5) \quad \|F_{\rho_\alpha^*(\zeta^N)} - F_{\pi^N}\|_\infty \leq \frac{1}{2N} + \left| \frac{\alpha - 1/2}{N} \right|.$$

Notice that the right-hand side of this expression is minimized for  $\alpha = 0.5$ . In fact, it is not difficult to check that the resampling scheme  $\rho_{1/2}^*$  is optimal in the sense that it minimises  $\|F_{\rho(\zeta^N)} - F_{\pi^N}\|_\infty$  among all resampling schemes  $\rho$ . One rationale for trying to minimize this quantity when considering deterministic resampling schemes is given by the generalized Koksma–Hlawka [Aistleitner and Dick (2015), Theorem 1] which implies that

$$(6) \quad |\rho(\zeta^N)(\varphi) - \pi^N(\varphi)| \leq V(\varphi) \|F_{\rho(\zeta^N)} - F_{\pi^N}\|_\infty$$

with  $V(\varphi)$  the variation of  $\varphi$  in  $\mathcal{X}$ .

We end this subsection by noting that inequality (5) shows that systematic resampling is consistent when applied on the ordered input particles  $\zeta^{N, \sigma_N^*}$ .

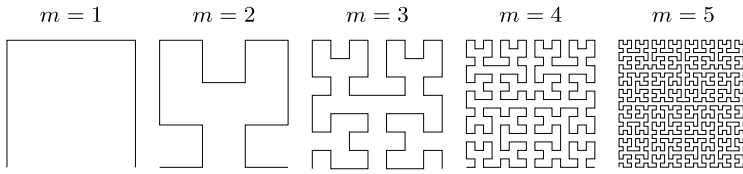


FIG. 1. *The Hilbert curve in dimension  $d = 2$  is defined as the limit of this sequence (source: Marc van Dongen).*

4.2. *Hilbert-ordered resampling schemes.* In this subsection, we generalize the results presented above to any dimension  $d \geq 1$ . The main challenge when  $d > 1$  is to find an ordering of particles  $\zeta^N$  which allows to improve upon the unordered version of the resampling scheme. Below we consider an ordering based on the Hilbert space filling curve.

4.2.1. *Hilbert space filling curve and related definitions.* For  $\pi, \pi' \in \mathcal{P}(\mathcal{X})$ , we use below the shorthand  $\|\pi - \pi'\|_\star = \|F_\pi - F_{\pi'}\|_\infty$ ; note that the “star” metric  $\|\cdot\|_\star$  is the multivariate generalization of the Kolmogorov metric. The star discrepancy of the point set  $u_{1:N}$  in  $[0, 1]^d$  is defined by

$$D_N^\star(u_{1:N}) = \left\| N^{-1} \sum_{i=1}^N \delta_{u_i} - \lambda_d \right\|_\star.$$

The Hilbert curve is a space-filling curve, that is a continuous surjective function  $H : [0, 1] \rightarrow [0, 1]^d$ . It is defined as the limit of the sequence of functions depicted (for  $d = 2$ ) in Figure 1. Precise details of the construction and some important properties of the Hilbert curve are given in Section S1.2 of the Supplementary Material [Gerber, Chopin and Whiteley (2019)]. In particular, the function  $H : [0, 1] \rightarrow [0, 1]^d$  is Hölder continuous with exponent  $1/d$  and is measure-preserving in the sense that  $\lambda_d(H(I)) = \lambda_1(I)$  for any measurable set  $I \in [0, 1]$ . This last property plays a crucial role in the derivation of the consistency results presented in the next subsection while the Hölder continuity of the Hilbert curve is central in our analysis of the variance of Hilbert-ordered stratified resampling (Section 4.2.3).

In the construction of the Hilbert curve, one is free to choose the value of  $H(0)$ , and we shall take it to be  $(0, \dots, 0)$ . The Hilbert curve admits a one-to-one Borel measurable pseudo-inverse  $h : [0, 1]^d \rightarrow [0, 1]$  such that  $H(h(x)) = x$  for all  $x \in [0, 1]^d$ , as shown in the next proposition.

PROPOSITION 2. *There exists a one-to-one Borel measurable function  $h : [0, 1]^d \rightarrow [0, 1]$  such that  $H(h(x)) = x$  for all  $x \in [0, 1]^d$ .*

For  $d = 1$ , we simply take  $H(x) = h(x) = x$  for  $x \in [0, 1]$ .

For a cubifiable set  $\mathcal{X}$  and diffeomorphism  $\psi \in \mathcal{D}(\mathcal{X})$ , we denote by  $h_{\mathcal{X},\psi}$  the one-to-one mapping  $x \mapsto h \circ \psi(x)$ . Remark that  $h_{\mathcal{X},\psi}(\mathcal{X}) = (0, 1)$  under the convention  $H(0) = (0, \dots, 0)$ . To simplify the notation in what follows, we associate to a cubifiable set  $\mathcal{X}$  a diffeomorphism  $\psi_{\mathcal{X}} \in \mathcal{D}(\mathcal{X})$  and use the shorthand  $h_{\mathcal{X}} = h_{\mathcal{X},\psi_{\mathcal{X}}}$ . In particular, when  $\mathcal{X} = (0, 1)^d$  we assume henceforth that  $\psi_{\mathcal{X}}(x) = x$  for all  $x \in \mathcal{X}$ .

We now define  $\sigma_N^*$  as a permutation of  $1 : N$  such that

$$h_{\mathcal{X}}(z_{\sigma_N^*(1)}) \leq \dots \leq h_{\mathcal{X}}(z_{\sigma_N^*(N)})$$

and use it to extend the definition of the ordered stratified resampling scheme  $\rho_{\text{strat}}^*$  introduced in the previous subsection to any  $d \geq 1$ ; that is, for any  $d \geq 1$  we define  $\rho_{\text{strat}}^* : [0, 1]^{\mathbb{N}} \times \mathcal{Z} \rightarrow \mathcal{P}_f(\mathcal{X})$  by

$$\rho_{\text{strat}}^*(u, z) = \rho_{\text{strat}}(u, (z_{\sigma_N^*(n)})_{n=1}^N), \quad (u, z) \in [0, 1]^{\mathbb{N}} \times \mathcal{Z}^N.$$

The resampling scheme  $\rho_{\text{strat}}^*(\zeta^N)$  is such that

$$(7) \quad X^{A^n} = \psi_{\mathcal{X}}^{-1} \circ H\left(F_{\pi_{h_{\mathcal{X}}}}^-\left(\frac{n-1+U_n}{N}\right)\right), \quad n \in 1 : N$$

and thus  $\rho_{\text{strat}}^*$  amounts to first sample from the empirical distribution  $\pi_{h_{\mathcal{X}}}^N$  using the stratified point set  $((n-1+U_n)/N)_{n=1}^N$  and then to “project” the resulting sample in the original set  $\mathcal{X}$  using the mapping  $\psi_{\mathcal{X}}^{-1} \circ H$ . Note that representation (7) of  $\rho_{\text{strat}}^*$  extends the one given in (4) for  $d = 1$  to any  $d \geq 1$ .

The ordered systematic resampling scheme  $\rho_{\text{sys}}^*$  is defined in a similar way.

Although this is not apparent from the notation, when  $d > 1$  the resampling schemes  $\rho_{\text{strat}}^*$  and  $\rho_{\text{sys}}^*$  depend on  $\psi_{\mathcal{X}}$  through  $\sigma_N^*$ , and, therefore, different choices for  $\psi_{\mathcal{X}}$  lead to different resampling mechanisms. Consequently, convergence results for these two resampling schemes will assume that the limiting distribution  $\pi$  on  $\mathcal{X}$  belongs to the subset  $\mathcal{P}_b^*(\mathcal{X})$  of  $\tilde{\mathcal{P}}_b(\mathcal{X})$  defined by  $\mathcal{P}_b^*(\mathcal{X}) = \{\pi \in \mathcal{P}_b(\mathcal{X}) : \pi_{\psi_{\mathcal{X}}} \in \mathcal{P}_b((0, 1)^d)\}$ .

To fix the ideas, when  $\mathcal{X} = \mathbb{R}^d$  one can take for  $\psi_{\mathcal{X}}$  the diffeomorphism  $\psi(x) = (\tilde{\psi}(x_1), \dots, \tilde{\psi}(x_d))$ , with  $\tilde{\psi} \in \mathcal{D}(\mathbb{R})$  defined by

$$\tilde{\psi}(x) = \frac{1}{2} + \frac{\sqrt{4+x^2}-2}{2x} \mathbb{1}_{\mathbb{R} \setminus \{0\}}(x), \quad x \in \mathbb{R}.$$

In this case, following Lemma 1, it is easily checked that  $\pi \in \mathcal{P}_b^*(\mathcal{X})$  when  $\pi \in \mathcal{P}_b(\mathcal{X})$  is such that  $\forall I \subseteq 1 : d$  and  $\forall x_{\setminus I} \in \times_{i \notin I} \mathcal{X}_i$  we have, for some  $C < +\infty$ ,  $\sup_{x_I \in \times_{i \in I} \mathcal{X}_i} p_{\pi}(x) \prod_{i \in I} |x_i|^2 \leq C$ .

4.2.2. *Consistency.* The following theorem provides a necessary and sufficient condition for the consistency of a resampling scheme.

**THEOREM 3.** *Let  $\mathcal{X}$  be a cubifiable set. Then a resampling scheme  $\rho$  is  $\tilde{\mathcal{P}}_b(\mathcal{X})$ -consistent if and only if, for any  $\pi \in \tilde{\mathcal{P}}_b(\mathcal{X})$  and sequence  $(\zeta^N)_{N \geq 1}$  such that  $\pi^N \xrightarrow{w} \pi$ ,  $\mathbb{P}$ -a.s., we have*

$$(8) \quad \lim_{N \rightarrow \infty} \|\rho(\zeta^N)_{h_{\mathcal{X}, \psi}} - \pi_{h_{\mathcal{X}, \psi}}^N\|_{\star} = 0, \quad \mathbb{P}\text{-a.s.}$$

for a  $\psi \in \mathcal{D}(\mathcal{X})$  such that  $\pi_{\psi} \in \mathcal{P}_b((0, 1)^d)$ .

This result is a consequence of Theorem 9 (see the [Appendix](#)) that establishes the equivalence between the weak convergence and the convergence in the sense of star metric, and shows that the Hilbert curve  $H$  and its pseudo-inverse  $h$  preserve these two modes of convergence.

A direct corollary of Theorem 3 is that any Hilbert-ordered resampling scheme satisfying the discrepancy condition in (10) below is consistent, and in particular the Hilbert-ordered versions of stratified and systematic resampling are consistent.

**COROLLARY 4.** *Let  $\mathcal{X}$  be a cubifiable set. For each  $N \geq 1$  and  $n \in 1 : N$ , let  $\phi_N^n : [0, 1]^N \rightarrow [0, 1]$  be a measurable function and consider a resampling scheme of the form*

$$(9) \quad a_N^n(u, \zeta_N) = F_N^{\sigma_N^*, -}(\phi_N^n(u))$$

with  $F_N^{\sigma_N^*, -}$  the inverse of the CDF  $F_N^{\sigma_N^*}(n) = \sum_{m=1}^N W^{\sigma_N(m)} \mathbb{1}(m \leq n)$ ,  $n \in 1 : N$ . Then a sufficient condition for such a resampling scheme to be  $\mathcal{P}_b^*(\mathcal{X})$ -consistent is that

$$(10) \quad \lim_{N \rightarrow +\infty} D_N^*(\phi_N^{1:N}(U)) = 0, \quad \mathbb{P}\text{-a.s.}$$

In particular,  $\rho_{\text{strat}}^*$  and  $\rho_{\text{sys}}^*$ , which correspond respectively to  $\phi_N^n(u) = (n - 1 + u_n)/N$  and  $\phi_N^n(u) = (n - 1 + u_1)/N$ , are  $\mathcal{P}_b^*(\mathcal{X})$ -consistent.

**4.2.3. Variance behaviour of Hilbert-ordered resampling.** The main goal of this subsection is to study in detail the convergence rate of the error variance for Hilbert-ordered stratified resampling.

The next result generalizes the first part of Theorem 2 to any  $d \geq 1$ .

**THEOREM 4.** *Let  $\mathcal{X}$  be a cubifiable set,  $\pi \in \mathcal{P}_b^*(\mathcal{X})$  have a strictly positive density, and let  $(\zeta^N)_{N \geq 1}$  be such that  $\pi^N \xrightarrow{w} \pi$ ,  $\mathbb{P}$ -a.s., and such that*

$$\lim_{N \rightarrow +\infty} \left( \max_{n \in 1:N} W^{n,N} \right) = 0, \quad \mathbb{P}\text{-a.s.}$$

Then, for any  $\varphi \in \mathcal{C}_b(\mathcal{X})$ ,

$$\text{Var}_{\zeta^N}[\rho_{\text{strat}}^*(\zeta^N)(\varphi)] = o(1/N), \quad \mathbb{P}\text{-a.s.}$$

Theorem 4 shows that under mild conditions Hilbert-ordered stratified resampling outperforms multinomial resampling asymptotically. The following result establishes its nonasymptotic behaviour under stronger assumptions on the test function  $\varphi$ .

**THEOREM 5.** *Let  $\mathcal{X}$  be a cubifiable set and  $\varphi : \mathcal{X} \rightarrow \mathbb{R}$  be a measurable function such that there exist constants  $C_{\varphi, \psi} < +\infty$  and  $\gamma \in (0, 1]$  verifying*

$$|\varphi \circ \psi_{\mathcal{X}}^{-1}(x) - \varphi \circ \psi_{\mathcal{X}}^{-1}(y)| \leq C_{\varphi, \psi_{\mathcal{X}}} \|x - y\|_2^\gamma, \quad \forall (x, y) \in (0, 1)^d.$$

Then for any  $N \geq 1$  we have

$$\text{Var}[\rho_{\text{strat}}^*(z)(\varphi)] \leq (2\sqrt{d+3})^{2\gamma} \frac{C_{\varphi, \psi_{\mathcal{X}}}^2}{N^{1+\frac{\gamma}{d}}}, \quad \forall z \in \mathcal{Z}^N.$$

The key tool to establish this result is the generalized Koksma–Hlawka inequality of Aistleitner and Dick [(2015), Theorem 1] that we already used in (6).

Note that, because of the use of the Hilbert curve in the resampling mechanism, the rate given in Theorem 5 cannot be improved by assuming differentiability on  $\varphi$ . This is true because the Hilbert curve is nowhere differentiable [see, e.g., Zumbusch (2003), Lemma 4.3, page 96]. We also note that the rate reported in Theorem 5 for  $\gamma = 1$  is in line with the one reported in He and Owen (2016), where for a random quadrature based on the Hilbert curve a variance of size  $\mathcal{O}(N^{-1-1/d})$  is found for a class of discontinuous functions having a Lipschitz component.

It should also be clear that the power  $1/d$  appearing in the upper bound of Theorem 5 arises because the Hilbert curve is Hölder continuous with exponent  $1/d$ . This latter is “optimal” in the sense that  $1/d$  is the best possible Hölder exponent for measure-preserving mappings from  $[0, 1]$  onto  $[0, 1]^d$  [Jaffard and Nicolay (2009), Lemma 6]. For this reason, it seems hard to improve the upper bound of Theorem 5 by considering an alternative ordering of the particles.

An interesting property of Theorem 5 is that it holds for any  $N \geq 1$  and requires no conditions on the weights and on the existence of a  $\pi \in \mathcal{P}(\mathcal{X})$  such that  $\pi^N \xrightarrow{w} \pi$ . At the same time, this suggests that the rate of  $N^{1+\gamma/d}$  is not optimal when a limiting distribution  $\pi$  exists. Indeed, Theorem 5 does not take into account that, in the definition of  $\rho_{\text{strat}}^*(\zeta^N)$  given in (7), the CDF  $F_{\pi_{h_{\mathcal{X}}}}^{\pi^N}$  may converge to  $F_{\pi_{h_{\mathcal{X}}}}$ , the CDF of  $\pi_{h_{\mathcal{X}}}$ , which is potentially a “smooth” function. This point is corrected in the next result.

**THEOREM 6.** *Consider the setup of Theorem 5, let  $(\zeta^N)_{N \geq 1}$  and  $\pi \in \mathcal{P}_b^*(\mathcal{X})$  be as in Theorem 4 and assume that*

$$(11) \quad \text{Var}_{\zeta^N} \left[ \frac{1}{N} \sum_{n=1}^N F_{\pi_{h_{\mathcal{X}}}}^- \left( \frac{n-1+U_n}{N} \right) \right] = o(N^{-2}), \quad \mathbb{P}\text{-a.s.}$$



Then, for any measurable function  $\varphi : \mathcal{X} \rightarrow \mathbb{R}$  satisfying the condition of Theorem 5, we have

$$(12) \quad \text{Var}_{\zeta^N}[\rho_{\text{strat}}^*(\zeta^N)(\varphi)] = o(N^{-(1+\frac{\gamma}{d})}), \quad \mathbb{P}\text{-a.s.}$$

When there exists a constant  $c > 0$  such that  $c^{-1}\lambda_d(A) \leq \pi(A) \leq c\lambda_d(A)$  for all measurable set  $A \subseteq \mathcal{X}$  condition (11) is verified.

We note that the rate in (12) does not only depend on the underlying rate in (11) but also on the speed at which  $\pi^N$  converges (in some sense) to  $\pi$ . More precisely, the rate in (12) depends on the rate at which the quantity  $v_N := \|F_{\pi_h \mathcal{X}}^-(u) - F_{\pi_h \mathcal{X}}^-(u)\|_\infty$  converges to 0 as  $N \rightarrow +\infty$ . In particular, under the extra assumptions of the second part of the theorem, the rate in (12) becomes  $\mathcal{O}(N^{-(1+\frac{2\gamma}{d})})$  when  $v_N = \mathcal{O}(1/N)$ .

**5. Implications for particle algorithms.** We apply in this section our previous results to the study of particle algorithms.

5.1. *Setup.* We consider a generic Feynman–Kac model, consisting of (a) a Markov chain, with initial distribution  $\mu(dx_0)$ , Markov kernels  $M_t : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{X})$ ,  $t \geq 1$ , acting from  $\mathcal{X}$  to itself; and (b) a sequence of measurable functions,  $G_0 : \mathcal{X} \rightarrow \mathbb{R}_+$ ,  $G_t : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_+$  for  $t \geq 1$ . The corresponding Feynman–Kac distributions are defined as

$$\mathbb{Q}_t(dx_{0:t}) = \frac{1}{L_t} \mu(dx_0) G_0(x_0) \prod_{s=1}^t M_t(x_{t-1}, dx_t) G_s(x_{s-1}, x_s),$$

where

$$L_t = \int_{\mathcal{X}^{t+1}} \mu(dx_0) G_0(x_0) \prod_{s=1}^t M_t(x_{t-1}, dx_t) G_s(x_{s-1}, x_s),$$

assuming  $L_t > 0$ . In practice, we are usually interested in approximating the so-called filtering distributions, that is, the marginal distributions  $\pi_t(dx_t) = \int_{x_{0:t-1} \in \mathcal{X}^t} \mathbb{Q}_t(dx_{0:t})$ . We also define  $\ell_t = L_t/L_{t-1} = (\mathbb{Q}_{t-1} M_t)(G_t)$  and the operators,  $V_0(\varphi) = \eta(\{\varphi - \eta(\varphi)\}^2)$ , and for  $t \geq 1$ ,

$$V_t(x_{t-1}, \varphi) = M_t(x_{t-1}, \{\varphi - M_t(\varphi)\}^2),$$

where  $M_t(x_{t-1}, \varphi) := \int_{\mathcal{X}} \varphi(x_t) M_t(x_{t-1}, dx_t)$ , and  $M_t(\varphi)$  is the function  $x_{t-1} \rightarrow M_t(x_{t-1}, \varphi)$ .

The subsequent results will rely on the following assumptions:

(G) Functions  $G_t$  are continuous and upper bounded.

(M) The Markov kernels  $M_t$  define a Feller process, that is,  $M_t(\varphi) \in \mathcal{C}_b(\mathcal{X})$  for all  $\varphi \in \mathcal{C}_b(\mathcal{X})$ .

---

**Algorithm 2** Standard particle filter

---

At time 0:

- (a) Generate (for  $n \in 1 : N$ )  $X_0^n \sim \mu(dx_0)$ .
- (b) Compute (for  $n \in 1 : N$ )  $w_0^n = G_0(X_0^n)$  and  $W_0^n = w_0^n / \sum_{m=1}^N w_0^m$ .

Recursively, for times  $t = 1, \dots, T$ :

- (a) Resample: for a given resampling scheme  $\rho$ , generate ancestor variables  $A_t^{1:N}$ , where  $A_t^n = a_N^n(U_t, \zeta_{t-1}^N)$ ,  $U_t \sim \mathbb{P}$ , and  $\zeta_{t-1}^N = (X_{t-1}^n, W_{t-1}^n)_{n=1}^N$  (as in Definition 3).
  - (b) Generate (for  $n \in 1 : N$ )  $X_t^n \sim M_t(X_{t-1}^{A_t^n}, dx_t)$ .
  - (c) Compute (for  $n \in 1 : N$ )  $w_t^n = G_t(X_{t-1}^{A_t^n}, X_t^n)$  and  $W_t^n = w_t^n / \sum_{m=1}^N w_t^m$ .
- 

A standard particle filter (Algorithm 2) generates at iteration  $t$  a weighted sample,  $(X_t^n, W_t^n)_{n=1}^N$ , which approximates  $\pi_t$  through the random measure  $\pi_t^N(dx_t) = \sum_{n=1}^N W_t^n \delta(X_t^n)$ .

5.2. *Consistency.* We first state an almost sure weak convergence result for Algorithm 2 under the condition that  $\rho$  is consistent for a suitable class of distributions [see Crisan (2001), Theorem 2.3.2, page 23, for a proof].

PROPOSITION 3. *Let  $\mathcal{P}_0 \subseteq \mathcal{P}(\mathcal{X})$  and assume that the Feynman–Kac model defined by  $(G_t)_{t \geq 0}$ ,  $\mu$  and  $(M_t)_{t \geq 1}$  is such that Assumptions (G) and (M) hold, and that  $\pi_t \in \mathcal{P}_0$  for all  $t \geq 0$ . Then, for any  $\mathcal{P}_0$ -consistent resampling scheme  $\rho : [0, 1]^N \times \mathcal{Z} \rightarrow \mathcal{P}_f(\mathcal{X})$  and  $t \geq 0$ , the particle approximation  $\pi_t^N := \sum_{n=1}^N W_t^n \delta(X_t^n)$  of  $\pi_t$  generated by Algorithm 2 is such that*

$$(13) \quad \pi_t^N \xrightarrow{w} \pi_t, \quad \mathbb{P}\text{-a.s.}$$

As a corollary, when  $\mathcal{X}$  is a cubifiable set and the assumptions of the proposition are satisfied with  $\mathcal{P}_0 = \tilde{\mathcal{P}}_b(\mathcal{X})$ , this result shows that Algorithm 2 based on stratified and SSP resampling is consistent in the sense that (13) holds for any  $t \geq 0$ .

We recall that (13) implies that, for any  $\varphi \in \mathcal{C}_b(\mathcal{X})$ ,  $\pi_t^N(\varphi) \rightarrow \pi_t(\varphi)$ ,  $\mathbb{P}$ -a.s. When stratified resampling is used in Algorithm 2 we note that, because this resampling mechanism dominates multinomial resampling in term of variance (see Section 2.2), it also holds true that  $\lim_{N \rightarrow +\infty} N \mathbb{E}[(\pi_t^N(\varphi) - \pi_t(\varphi))^2] < +\infty$  for any  $\varphi \in \mathcal{C}_b(\mathcal{X})$ . For unbounded measurable function  $\varphi : \mathcal{X} \rightarrow \mathbb{R}$  such that  $\pi_t(\varphi) < +\infty$ , the results in Cappé, Moulines and Rydén [(2005), Chapter 9] imply that  $\pi_t^N(\varphi) \rightarrow \pi_t(\varphi)$  in  $\mathbb{P}$ -probability.

5.3. *Central limit theorem.* As shown in the previous section, the “noise” introduced by the Hilbert ordered stratified resampling scheme  $\rho_{\text{strat}}^*$  converges to zero faster than the usual  $\mathcal{O}(N^{-1})$  Monte Carlo rate. The next result formalises the intuitive idea that, when Algorithm 2 is based on this resampling mechanism, the resampling step does not contribute to the asymptotic variance of the quantity  $N^{1/2}(\pi_t^N(\varphi) - \pi_t(\varphi))$ . For sake of completeness, Theorem 7 also presents results for the multinomial resampling ( $\rho_{\text{multi}}$ ) and residual reampling ( $\rho_{\text{res/multi}}$ ) schemes for which a central limit theorem also exists [see Chopin (2004), Künsch (2005), Douc, Cappé and Moulines (2005)].

**THEOREM 7.** *For Algorithm 2, assuming that  $\mathcal{X}$  is a cubifiable set,  $\pi_t \in \mathcal{P}_b^*(\mathcal{X})$  for all  $t \geq 0$ ,  $\rho \in \{\rho_{\text{multi}}, \rho_{\text{res/multi}}, \rho_{\text{strat}}^*\}$  and that the Feynman–Kac model fulfils assumptions (G) and (M), for any test function  $\varphi \in \mathcal{C}_b(\mathcal{X})$  we have that (for any  $t \geq 0$ )*

$$N^{1/2} \left\{ \sum_{n=1}^N W_t^n \varphi(X_t^n) - \pi_t(\varphi) \right\} \xrightarrow{w} \mathcal{N}_d(0, \mathcal{V}_t[\varphi]),$$

where the  $\mathcal{V}_t(\varphi)$  are defined recursively as follows:  $\tilde{\mathcal{V}}_0[\varphi] = V_0(\varphi)$ ,

$$\mathcal{V}_t[\varphi] = \frac{1}{\varrho_t^2} \tilde{\mathcal{V}}_t[G_t\{\varphi - \pi_t(\varphi)\}],$$

$$\widehat{\mathcal{V}}_t[\varphi] = \mathcal{V}_t[\varphi] + R_t(\rho, \varphi),$$

$$\tilde{\mathcal{V}}_{t+1}[\varphi] = \widehat{\mathcal{V}}_t[M_{t+1}(\varphi)] + \pi_t[V_{t+1}(\varphi)]$$

and

$$0 = R_t(\rho_{\text{strat}}^*, \varphi) \leq R_t(\rho_{\text{res/multi}}, \varphi) \leq R_t(\rho_{\text{multi}}, \varphi).$$

The proof is a simple combination of Theorem 4 and the proofs in the aforementioned papers (see the Supplementary Material [Gerber, Chopin and Whiteley (2019)]).

An obvious corollary of this theorem is that ordered stratified resampling dominates multinomial and residual resampling, in terms of the asymptotic variance of particle estimates generated by a particle filter. In fact, since the contribution of the resampling step is zero when ordered stratified resampling is used, this particular scheme may be declared as optimal (again, relative to the asymptotic variance for any test function).

5.4. *A note on the auxiliary particle filter.* The auxiliary particle filter [APF, Pitt and Shephard (1999)] is a variation on the standard particle filter, where the resampling weights are “twisted” using some function  $\eta_t : \mathcal{X} \rightarrow \mathbb{R}_{>0}$ ; that is, the resampling weight of ancestor  $X_{t-1}^m$  is  $\tilde{W}_t^m \propto W_{t-1}^m \times \eta_{t-1}(X_{t-1}^m)$ ;  $\sum_{n=1}^N \tilde{W}_{t-1}^n = 1$ .

When a particle  $X_t^n$  originates from ancestor  $X_{t-1}^m$ , that is,  $A_t^n = m$ , it is assigned (un-normalised) weight  $w_t^n = G_t(X_{t-1}^m, X_t^n)W_{t-1}^m/\widetilde{W}_{t-1}^m$ , so as to correct for the discrepancy between the resampling weights and the actual weights.

Of particular interest is particle estimate

$$\ell_t^N = \frac{1}{N} \sum_{n=1}^N w_t^n = \frac{1}{N} \sum_{n=1}^N \frac{W_{t-1}^{A_t^n}}{\widetilde{W}_{t-1}^{A_t^n}} G_t(X_{t-1}^{A_t^n}, X_t^n)$$

of normalising constant  $\ell_t$ , and the cumulative product  $L_t^N = \prod_{s=0}^t \ell_t^N$ , which estimates  $L_t = \prod_{s=0}^t \ell_t$ . The latter quantity usually corresponds to the likelihood of the data observed up to time  $t$  (for a certain model) and thus plays a central role in parameter estimation methods [e.g., particle Markov chain Monte Carlo, [Andrieu, Doucet and Holenstein \(2010\)](#)].

**THEOREM 8.** *Consider the APF Algorithm (as described above), a given Feynman–Kac model such that Assumptions (G) and (M) hold, and assume that functions  $\eta_0, \dots, \eta_{t-2}$  are fixed. For  $\rho = \rho_{\text{multi}}$ , the function  $\eta_{t-1}(x_{t-1}) = \sqrt{M_t(x_{t-1}, G_t^2)}$  minimises the variance of particle estimates  $\ell_t^N$  and  $L_t^N$ .*

*For  $\rho = \rho_{\text{strat}}^*$ , assuming in addition that  $\mathcal{X}$  is a compact cubifiable set, the quantities  $N \text{Var}[\ell_t^N]$  and  $N \text{Var}[L_t^N]$  converge to a limit which is minimal for  $\eta_{t-1} = \eta_{t-1}^*$ , where  $\eta_{t-1}^*(x_{t-1}) = \sqrt{V_t(x_{t-1}, G_t)}$ , among functions  $\eta_{t-1} \in \mathcal{C}_b(\mathcal{X})$  that are positive almost everywhere. (In particular,  $\eta_{t-1}^*$  itself is assumed to be positive everywhere.)*

The usual recommendation [e.g., [Johansen and Doucet \(2008\)](#)] is to take  $\eta_{t-1}(x_{t-1}) = M_t(x_{t-1}, G_t)$  (or some approximation of this quantity). Under multinomial resampling, and in the “perfectly adapted” case (where  $G_t$  depends only on  $x_{t-1}$ ), the proposition above shows that this choice is indeed optimal. Unfortunately, it also shows that the choice of the auxiliary function in the APF should actually depend on the resampling scheme. This point deserves further study, which we leave for future research. We refer to [Douc, Moulines and Olsson \(2009\)](#) for related results on optimal auxiliary functions (relative to the asymptotic variance for a given test function) and [Cornebise, Moulines and Olsson \(2008\)](#) for some numerical scheme to approximate these optimal auxiliary functions within a parametric family. But again both papers assume multinomial resampling, and their results and proposed methodology should be adapted if another resampling scheme is used.

**5.5. Numerical experiments.** We compare in this section the approximation  $(\pi_t^N)_{t=0}^T$  of  $(\pi_t)_{t=0}^T$  generated by Algorithm 2 under the resampling schemes  $\rho_{\text{strat}}$  (stratified resampling),  $\rho_{\text{strat}}^*$  (ordered stratified resampling) and  $\rho_{\text{ssp}}$  (SSP resampling).

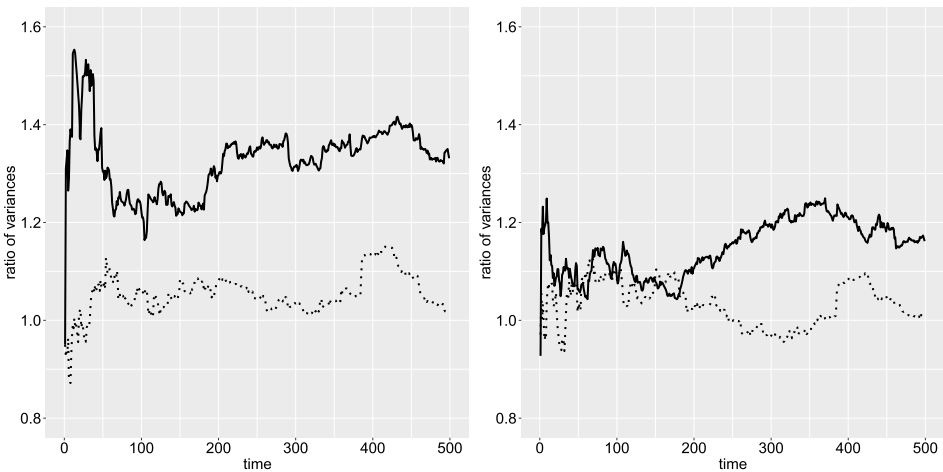
Following [Guarniero, Johansen and Lee \(2017\)](#), we consider the linear Gaussian state-space models where  $X_0 \sim \mathcal{N}_d(0, I_d)$ , and, for  $t = 1, \dots, T$ ,

$$\begin{aligned} X_t &= F X_{t-1} + V_t, & V_t &\sim \mathcal{N}_d(0, I_d), \\ Y_t &= X_t + W_t, & W_t &\sim \mathcal{N}_d(0, I_d), \end{aligned}$$

with  $F = (\alpha^{|i-j|+1})_{i,j=1}^d$ ,  $\alpha = 0.4$ ,  $T = 500$  and  $d = 5$ . We focus on the problem of estimating the log-likelihood of the model,  $\log p(y_{1:T})$ , which is estimated from the output of [Algorithm 2](#) by  $\log L_T^N = \sum_{t=0}^T \log \ell_t^N$  (see [Section 5.4](#)).

We consider two Feynman–Kac models; a “bootstrap” model, where the Markov kernel  $M_t(x_{t-1}, dx_t)$  corresponds to the law of  $X_t | X_{t-1} = x_{t-1}$ ,  $G_t(x_{t-1}, x_t)$  is the probability density of  $Y_t | X_t = x_t$ ; and a “guided” model, where  $M_t(x_{t-1}, dx_t)$  is the Gaussian distribution  $\mathcal{N}_d((y_t + Fx_{t-1})/2, I_d/2)$ ,  $G_t(x_{t-1}, x_t)$  is the probability density of  $\mathcal{N}_d(Fx_{t-1}, 2I_d)$  at point  $y_t$ . Both Feynman–Kac formalisms are such that  $\pi_t$  is the filtering distribution at time  $t$  of the model above. The point of the guided formalism is to reduce the variance of the weights (at each time  $t$ ), and thus to reduce the variance of particle estimates.

[Figure 2](#) shows the variance of the estimator  $\log L_T^N$  obtained under the two above Feynman–Kac formalisms, as a function of  $t \in 1 : T$ , and for the resampling schemes  $\rho_{\text{strat}}$ ,  $\rho_{\text{strat}}^*$  and  $\rho_{\text{spp}}$ . For each resampling scheme, the results of [Figure 2](#) are based on 1000 independent runs of the two particle algorithms we are considering, with  $N = 2^{13}$  particles.



**FIG. 2.** Estimation of the log-likelihood function as a function of  $t$ . The left (resp., right) plot gives the variance of SMC based on unordered stratified resampling divided by that of SMC based on Hilbert-ordered stratified resampling (resp. unordered SSP resampling). Continuous lines are for SMC based on the guided proposal while the dotted line is for the bootstrap particle filter. Results are based on 1000 independent runs of the algorithms with  $N = 2^{13}$  particles.

As expected from the results of Section 4, the variance of  $\log L_t^N$  is smaller with  $\rho_{\text{strat}}^*$  than with  $\rho_{\text{strat}}$ ; the relative gains are larger when the guided formalism is used (where the variances under  $\rho_{\text{strat}}$  are about 40% higher than under  $\rho_{\text{strat}}^*$ ). The results presented in Figure 2 suggest that  $\rho_{\text{SSP}}$  is preferable to  $\rho_{\text{strat}}$ . This is particularly true with the guided formalism where the variances under  $\rho_{\text{strat}}$  are about 20% higher than when  $\rho_{\text{SSP}}$  is used. Lastly, the variances under SSP resampling are larger than under ordered stratified resampling but  $\rho_{\text{SSP}}$  has the advantage to be faster. Indeed, SSP resampling requires  $\mathcal{O}(N)$  operations against  $\mathcal{O}(N \log N)$  for  $\rho_{\text{strat}}^*$ .

**6. Conclusion.** Our results support the practice in the SMC literature to abandon multinomial resampling for stratified resampling by providing strong theoretical guarantees for this resampling scheme, which has the remarkable property to be both cheaper and more accurate than multinomial resampling. For the same reasons, our results should encourage practitioners to abandon residual resampling for a version of this residual method where the multinomial resampling step is replaced by a stratified resampling step.

The systematic resampling scheme fails to produce offspring numbers that are negatively associated. As an alternative to it we have introduced the SSP resampling algorithm which (1) is similar to systematic resampling in term of offspring numbers and (2) verifies the conditions of our general consistency result. We also built an example suggesting that any general consistency results for systematic resampling would require to take into account the order of the input particles and have established its validity when they are ordered along the Hilbert curve.

Our practical recommendation is to prefer SSP resampling to systematic resampling since both have similar properties while only the former has been proven to be consistent. Systematic resampling has the advantage to be faster than SSP resampling but in most cases this gain is likely to be imperceptible. Our simulation study suggests that SSP resampling outperforms also stratified resampling in term of variance but no theoretical result exists to support this observation.

We have also derived various results showing that the variance of stratified resampling goes to zero faster than  $N^{-1}$  when applied on an input point set ordered along the Hilbert curve, and notably a nonasymptotic bound of size  $N^{-1-\frac{1}{d}}$ . Unsurprisingly, when the dimension of the state-space is small and/or when a good proposal distribution is available, our simulation results show that ordering the particle before applying stratified resampling may lead to important variance reduction. These theoretical results on the variance of Hilbert ordered stratified resampling are also of particular interest for sequential quasi-Monte Carlo [Gerber and Chopin (2015)], a quasi-Monte Carlo version of SMC, that converges at a faster but currently unknown rate.

APPENDIX: CONVERGENT SEQUENCES OF PROBABILITY MEASURES:  
STAR NORM AND TRANSFORMATIONS THROUGH THE HILBERT  
CURVE AND ITS INVERSE

The following theorem is the main tool for establishing Theorem 3.

**THEOREM 9.** *Let  $\mathcal{X}$  be a cubifiable set,  $(\pi^N)_{N \geq 1}$  be a sequence in  $\mathcal{P}(\mathcal{X})$ ,  $\pi \in \tilde{\mathcal{P}}_b(\mathcal{X})$  and  $\psi \in \mathcal{D}(\mathcal{X})$  be such that  $\pi_\psi \in \mathcal{P}_b((0, 1)^d)$ . Then the following assertions are equivalent:*

- (i)  $\pi^N \xrightarrow{w} \pi$ ;
- (ii)  $\lim_{N \rightarrow +\infty} \|\pi^N - \pi\|_\star = 0$ ;
- (iii)  $\lim_{N \rightarrow +\infty} \|\pi_{h_{\mathcal{X}, \psi}}^N - \pi_{h_{\mathcal{X}, \psi}}\|_\star = 0$ ;
- (iv)  $\pi_{h_{\mathcal{X}, \psi}}^N \xrightarrow{w} \pi_{h_{\mathcal{X}, \psi}}$ .

Implications (ii)  $\Rightarrow$  (iii) and (iii)  $\Rightarrow$  (ii), respectively, are due to Gerber and Chopin [(2015), Theorem 3] and Schretter et al. [(2016), Theorem 1]. Implications (ii)  $\Rightarrow$  (i) and (iii)  $\Rightarrow$  (iv) are direct applications of the Portmanteau lemma [e.g., van der Vaart (1998), Lemma 2.2, page 6]. Implication (i)  $\Rightarrow$  (ii) for  $d = 1$  holds by Polyà’s theorem [Pólya (1920); see also Bickel and Millar (1992), result (A.1)]; note that Polyà’s theorem only requires that  $\pi \in \mathcal{P}(\mathcal{X})$  is such that  $F_\pi$  is continuous. Implication (i)  $\Rightarrow$  (ii) for  $d > 1$  is new and proved following a similar argument as in Kuipers and Niederreiter [(1974), Theorem 1.2, page 89] while implication (iv)  $\Rightarrow$  (iii) is a consequence of Polyà’s theorem and of the continuity of  $F_{\pi_{h_{\mathcal{X}, \psi}}}$ , which is established in the next lemma.

**LEMMA 2.** *Let  $\mathcal{X}$  be a cubifiable set,  $\pi \in \tilde{\mathcal{P}}_b(\mathcal{X})$  and  $\psi \in \mathcal{D}(\mathcal{X})$  be such that  $\pi_\psi \in \mathcal{P}_b((0, 1)^d)$ . Then  $\pi_{h_{\mathcal{X}, \psi}}$  is a continuous probability measure on  $(0, 1)$ .*

We also note the proofs of implications (ii)  $\Rightarrow$  (iii) and (iii)  $\Rightarrow$  (ii) in Gerber and Chopin (2015), Gerber and Chopin (2017), Schretter et al. (2016) implicitly assume that the sequence  $(\pi^N)_{N \geq 1}$  is such that [with  $\mathcal{X} = (0, 1)^d$ ]

$$\pi^N(\mathcal{H}_d) = 0 \quad \text{for all } N \text{ large enough,}$$

where  $\mathcal{H}_d$  is the set of points of  $[0, 1]^d$  that have more than pre-image through  $H$ . This point is corrected in the Supplementary Material [Gerber, Chopin and Whiteley (2019)] where a complete proof of Theorem 9 is provided.

**Acknowledgements.** We thank Patrice Bertail, Anthony Lee and Matthieu Wihelm for useful comments.

SUPPLEMENTARY MATERIAL

**Supplementary material: Proofs** (DOI: [10.1214/18-AOS1746SUPP](https://doi.org/10.1214/18-AOS1746SUPP); .pdf). This supplement contains all the proofs.

## REFERENCES

- AISTLEITNER, C. and DICK, J. (2015). Functions of bounded variation, signed measures, and a general Koksma–Hlawka inequality. *Acta Arith.* **167** 143–171. [MR3312093](#)
- ANDRIEU, C., DOUCET, A. and HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 269–342. [MR2758115](#)
- BARBE, P. and BERTAIL, P. (1995). *The Weighted Bootstrap. Lecture Notes in Statistics* **98**. Springer, New York. [MR2195545](#)
- BERTAIL, P., CHAUTRU, E. and CLÉMENÇON, S. (2017). Empirical processes in survey sampling with (conditional) Poisson designs. *Scand. J. Stat.* **44** 97–111. [MR3619696](#)
- BICKEL, P. J. and MILLAR, P. W. (1992). Uniform convergence of probability measures on classes of functions. *Statist. Sinica* **2** 1–15. [MR1152295](#)
- BROWN KRAMER, J., CUTLER, J. and RADCLIFFE, A. J. (2011). Negative dependence and Srinivasan’s sampling process. *Combin. Probab. Comput.* **20** 347–361. [MR2784632](#)
- CAPPÉ, O., MOULINES, E. and RYDÉN, T. (2005). *Inference in Hidden Markov Models*. Springer, New York. [MR2159833](#)
- CARPENTER, J., CLIFFORD, P. and FEARNHEAD, P. (1999). Improved particle filter for nonlinear problems. *IEE Proc. Radar Sonar Navig.* **146** 2–7.
- CHOPIN, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.* **32** 2385–2411. [MR2153989](#)
- CORNEBISE, J., MOULINES, É. and OLSSON, J. (2008). Adaptive methods for sequential importance sampling with application to state space models. *Stat. Comput.* **18** 461–480. [MR2461889](#)
- CRISAN, D. (2001). Particle filters—A theoretical perspective. In *Sequential Monte Carlo Methods in Practice* (A. Doucet, N. de Freitas and N. J. Gordon, eds.) 17–41. Springer, New York. [MR1847785](#)
- CRISAN, D. and DOUCET, A. (2002). A survey of convergence results on particle filtering methods for practitioners. *IEEE Trans. Signal Process.* **50** 736–746. [MR1895071](#)
- DEL MORAL, P. and DOUCET, A. (2003). On a class of genealogical and interacting Metropolis models. In *Séminaire de Probabilités XXXVII. Lecture Notes in Math.* **1832** 415–446. Springer, Berlin. [MR2053058](#)
- DEVILLE, J.-C. and TILLÉ, Y. (1998). Unequal probability sampling without replacement through a splitting method. *Biometrika* **85** 89–101. [MR1627234](#)
- DOUC, R., CAPPÉ, O. and MOULINES, E. (2005). Comparison of resampling schemes for particle filtering. In *ISPA 2005. Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis* 64–69. IEEE, New York.
- DOUC, R., MOULINES, É. and OLSSON, J. (2009). Optimality of the auxiliary particle filter. *Probab. Math. Statist.* **29** 1–28. [MR2552996](#)
- DOUCET, A., DE FREITAS, N. and GORDON, N. J. (2001). *Sequential Monte Carlo Methods in Practice*. Springer, New York.
- DUBHASHI, D., JONASSON, J. and RANJAN, D. (2007). Positive influence and negative dependence. *Combin. Probab. Comput.* **16** 29–41. [MR2286510](#)
- DUBHASHI, D. and RANJAN, D. (1998). Balls and bins: A study in negative dependence. *Random Structures Algorithms* **13** 99–124. [MR1642566](#)
- GERBER, M. and CHOPIN, N. (2015). Sequential quasi Monte Carlo. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 509–579. [MR3351446](#)
- GERBER, M. and CHOPIN, N. (2017). Convergence of sequential quasi-Monte Carlo smoothing algorithms. *Bernoulli* **23** 2951–2987. [MR3654796](#)
- GERBER, M., CHOPIN, N. and WHITELEY, N. (2019). Supplement to “Negative association, ordering and convergence of resampling methods.” DOI:10.1214/18-AOS1746SUPP.
- GORDON, N. J., SALMOND, D. J. and SMITH, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc. F, Commun. Radar Signal Process.* **140** 107–113.



- GUARNIERO, P., JOHANSEN, A. M. and LEE, A. (2017). The iterated auxiliary particle filter. *J. Amer. Statist. Assoc.* **112** 1636–1647. [MR3750887](#)
- HE, Z. and OWEN, A. B. (2016). Extensible grids: Uniform sampling on a space filling curve. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 917–931. [MR3534356](#)
- JAFFARD, S. and NICOLAY, S. (2009). Pointwise smoothness of space-filling functions. *Appl. Comput. Harmon. Anal.* **26** 181–199. [MR2490213](#)
- JOAG-DEV, K. and PROSCHAN, F. (1983). Negative association of random variables, with applications. *Ann. Statist.* **11** 286–295. [MR0684886](#)
- JOHANSEN, A. M. and DOUCET, A. (2008). A note on auxiliary particle filters. *Statist. Probab. Lett.* **78** 1498–1504. [MR2528342](#)
- KITAGAWA, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comput. Graph. Statist.* **5** 1–25. [MR1380850](#)
- KUIPERS, L. and NIEDERREITER, H. (1974). *Uniform Distribution of Sequences: Pure and Applied Mathematics*. Wiley-Interscience [John Wiley & Sons], New York–London–Sydney. [MR0419394](#)
- KÜNSCH, H. R. (2005). Recursive Monte Carlo filters: Algorithms and theoretical analysis. *Ann. Statist.* **33** 1983–2021. [MR2211077](#)
- L’ECUYER, P. and LEMIEUX, C. (2000). Variance reduction via lattice rules. *Manage. Sci.* **46** 1214–1235.
- LIU, J. S. and CHEN, R. (1998). Sequential Monte Carlo methods for dynamic systems. *J. Amer. Statist. Assoc.* **93** 1032–1044. [MR1649198](#)
- MADOW, W. G. and MADOW, L. H. (1944). On the theory of systematic sampling. I. *Ann. Math. Stat.* **15** 1–24. [MR0009836](#)
- NIEDERREITER, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods*. CBMS-NSF Regional Conference Series in Applied Mathematics **63**. SIAM, Philadelphia, PA. [MR1172997](#)
- PITT, M. K. and SHEPHARD, N. (1999). Filtering via simulation: Auxiliary particle filters. *J. Amer. Statist. Assoc.* **94** 590–599. [MR1702328](#)
- PÓLYA, G. (1920). Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem. *Math. Z.* **8** 171–181. [MR1544437](#)
- SCHRETTTER, C., HE, Z., GERBER, M., CHOPIN, N. and NIEDERREITER, H. (2016). Van der Corput and golden ratio sequences along the Hilbert space-filling curve. In *Monte Carlo and Quasi-Monte Carlo Methods*. Springer Proc. Math. Stat. **163** 531–544. Springer, Cham. [MR3536680](#)
- SHAO, Q.-M. (2000). A comparison theorem on moment inequalities between negatively associated and independent random variables. *J. Theoret. Probab.* **13** 343–356. [MR1777538](#)
- SRINIVASAN, A. (2001). Distributions on level-sets with applications to approximation algorithms. In *42nd IEEE Symposium on Foundations of Computer Science (Las Vegas, NV, 2001)* 588–597. IEEE Computer Soc., Los Alamitos, CA. [MR1948748](#)
- STEIN, M. (1987). Large sample properties of simulations using Latin hypercube sampling. *Technometrics* **29** 143–151. [MR0887702](#)
- TILLÉ, Y. (2006). *Sampling Algorithms*. Springer, New York. [MR2225036](#)
- VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics **3**. Cambridge Univ. Press, Cambridge. [MR1652247](#)
- ZUMBUSCH, G. (2003). *Parallel Multilevel Methods*. Springer, Berlin.

M. GERBER  
 N. WHITELEY  
 SCHOOL OF MATHEMATICS  
 UNIVERSITY OF BRISTOL  
 UNIVERSITY WALK  
 BRISTOL BS8 1TW  
 UNITED KINGDOM  
 E-MAIL: [mathieu.gerber@bristol.ac.uk](mailto:mathieu.gerber@bristol.ac.uk)  
[nick.whiteley@bristol.ac.uk](mailto:nick.whiteley@bristol.ac.uk)

N. CHOPIN  
 CREST-ENSAE  
 3, AVENUE PIERRE LAROUSSE  
 92245 MALAKOFF CEDEX  
 FRANCE  
 E-MAIL: [nicolas.chopin@ensae.fr](mailto:nicolas.chopin@ensae.fr)